**人工智能训练师（三级）操作技能考核**

**试题单**

准考证号：

试题代码：

试题名称：金融机构信用评估系统中的业务数据审核流程设计

考核时间：30min

**1.**场地设备要求

人工智能训练师主机：CPU（intel i5 及以上）、内存（不少于 16GB）、操作系统（windows10）、支持深度学习训练；

**2.**工作任务

某金融机构计划引入智能信用评估系统，通过分析客户的历史交易数据和信用记录，使用机器学习算法预测客户的信用风险等级，从而辅助贷款审批和风险控制。为了确保数据的准确性和可靠性，该机构需要设计并实现一套全面的业务数据审核流程，确保数据在进入信用评估系统之前经过严格的审核和清洗。

我们提供一个客户信用数据集（credit\_data.csv），包含以下字段：

* CustomerID: 客户ID
* Name: 客户姓名
* Age: 年龄
* Income: 收入
* LoanAmount: 贷款金额
* LoanTerm: 贷款期限（月）
* CreditScore: 信用评分
* Default: 是否违约（0: 否，1: 是）
* TransactionHistory: 历史交易记录（JSON格式）

你作为人工智能训练师，根据提供的credit\_data.csv数据集和Python代码框架（1.1.3.ipynb），完成以下数据的审核和处理任务，确保数据的准确性和可靠性。请按照以下要求完成任务，确保结果准确并保存相应的截图。

（1）数据完整性审核：

通过运行Python代码（1.1.3.ipynb）检查数据集中的每个字段是否存在缺失值和重复值。将上述审核结果截图以JPG的格式保存，命名为“1.1.3-1”。

（2）数据合理性审核：

通过运行Python代码（1.1.3.ipynb）审核以下字段的合理性：

* 年龄：应在18到70岁之间。
* 收入：应大于2000。
* 贷款金额：应小于收入的5倍。
* 信用评分：应在300到850之间。

对不合理的数据进行标记，并将审核结果截图以JPG的格式保存，命名为“1.1.3-2”。

（3）通过运行Python代码（1.1.3.ipynb）对数据进行清洗，处理异常值。具体要求如下：

* 将不合理的数据进行标记，并对异常值所在行进行删除；
* 清洗后的数据保存为新文件cleaned\_credit\_data.csv。

所有结果文件储存在桌面新建的考生文件夹中，文件夹命名为“准考证号+身份证号后六位”。

**3.**技能要求

（1）能结合人工智能技术要求和业务特征，设计整套业务数据处理流程；

（2）能结合人工智能技术要求和业务特征，设计整套业务数据审核流程

**4.**质量指标

（1）完整性指标：数据集中无缺失值和重复记录。

（2）合理性指标：所有数据点符合业务规则，无异常值存在。

（3）清洗效果指标：清洗后的数据集完整、合理，且适于建模分析。

**人工智能训练师（三级）操作技能考核**

**试题评分表**

准考证号：

试题代码：

试题名称：金融机构信用评估系统中的业务数据审核流程设计

考核时间：30min

测量分评分表

| 细则编号 | 配分 | 评分细则描述 | 规定或  标称值 | 结果或  实际值 | 得分 |
| --- | --- | --- | --- | --- | --- |
| M1 | 2 | 数据缺失值统计代码正确得2分 | 根据数据 |  |  |
| M2 | 2 | 数据重复值统计代码正确得2分 | 根据数据 |  |  |
| M3 | 1 | 1.1.3-1.jpg中CustomerID缺失值统计正确：得1分； | 根据数据 |  |  |
| M4 | 1 | 1.1.3-1.jpg中Name缺失值统计正确：得1分； | 根据数据 |  |  |
| M5 | 1 | 1.1.3-1.jpg中Age缺失值统计正确：得1分； | 根据数据 |  |  |
| M6 | 1 | 1.1.3-1.jpg中Income缺失值统计正确：得1分； | 根据数据 |  |  |
| M7 | 1 | 1.1.3-1.jpg中LoanAmount缺失值统计正确：得1分； | 根据数据 |  |  |
| M8 | 1 | 1.1.3-1.jpg中LoanTerm缺失值统计正确：得1分； | 根据数据 |  |  |
| M9 | 1 | 1.1.3-1.jpg中CreditScore缺失值统计正确：得1分； | 根据数据 |  |  |
| M10 | 1 | 1.1.3-1.jpg中Default缺失值统计正确：得1分； | 根据数据 |  |  |
| M11 | 1 | 1.1.3-1.jpg中TransactionHistory缺失值统计正确：得1分； | 根据数据 |  |  |
| M12 | 2 | Age数据的合理性审核代码正确得2分； | 根据数据 |  |  |
| M13 | 2 | Income数据的合理性审核代码正确得2分； | 根据数据 |  |  |
| M14 | 2 | LoanAmount数据的合理性审核代码正确得2分； | 根据数据 |  |  |
| M15 | 2 | CreditScore数据的合理性审核代码正确得2分； | 根据数据 |  |  |
| M16 | 1 | 数据合理性检查结果(1.1.3-2.jpg)正确：得1分； | 根据数据 |  |  |
| M17 | 2 | 保存清洗后的代码正确得2分 | 根据数据 |  |  |
| M18 | 1 | 保存cleaned\_credit\_data.csv文件成功：得1分； | 根据数据 |  |  |
| 合计配分 | 25 | 合计得分 | |  |  |

**参考答案**：

Python参考代码（1.1.3.ipynb）：

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

# 读取数据集

data = pd.read\_csv('credit\_data.csv')

# 1. 数据完整性审核

missing\_values = data.isna().sum() # 数据缺失值统计

duplicate\_values = data.duplicated().sum() # 数据重复值统计

# 输出结果

print("缺失值统计:")

print(missing\_values)

print("重复值统计:")

print(duplicate\_values)

# 2. 数据合理性审核

data['is\_age\_valid'] = data['Age'].between(18, 70) # Age数据的合理性审核

data['is\_income\_valid'] = data['Income'] > 2000 # Income数据的合理性审核

data['is\_loan\_amount\_valid'] = data['LoanAmount'] < (data['Income'] \* 5) # LoanAmount数据的合理性审核

data['is\_credit\_score\_valid'] = data['CreditScore'].between(300, 850) # CreditScore数据的合理性审核

# 合理性检查结果

validity\_checks = data[['is\_age\_valid', 'is\_income\_valid', 'is\_loan\_amount\_valid', 'is\_credit\_score\_valid']].all(axis=1)

data['is\_valid'] = validity\_checks

# 输出结果

print("数据合理性检查:")

print(data[['is\_age\_valid', 'is\_income\_valid', 'is\_loan\_amount\_valid', 'is\_credit\_score\_valid', 'is\_valid']].describe())

# 3. 数据清洗和异常值处理

# 标记不合理数据

invalid\_rows = data[~data['is\_valid']]

# 删除不合理数据行

cleaned\_data = data[data['is\_valid']]

# 删除标记列

cleaned\_data = cleaned\_data.drop(columns=['is\_age\_valid', 'is\_income\_valid', 'is\_loan\_amount\_valid', 'is\_credit\_score\_valid', 'is\_valid'])

# 保存清洗后的数据

cleaned\_data.to\_csv('cleaned\_credit\_data.csv', index=False)

print("数据清洗完成，已保存为 'cleaned\_credit\_data.csv'")

1.1.3-1.jpg



1.1.3-2.jpg

